THE ADVANTAGES IN USING ORTHOGONALISED TERMS IN A POLYNOMIAL FOR CURVE-FITTING

By V. SATAKOPAN

(Received for publication, December 6, 1941)

ABSTRACT. The paper briefly discusses the relationships between the least squares solutions of Polynomial Constants obtained by using :—

(1) ordinary power terms (solutions discussed by S. M. Kerawala (1941) and

(2) orthogonalised terms (R. A. Fisher's method),

when the polynomial is fitted to a series of observations.

It is shown how all the constants of polynomials up to the rth degree obtained by using simple power terms can be determined from the $(r+1)$ constants obtained from the orthogonal terms.

The advantages of using the orthogonalised terms instead of ordinary power terms are also indicated.

S. M. Kerawala (1941), in an interesting paper which appears in a recent issue of the Indian Journal of Physics has discussed the least squares solutions of the constants of polynomials up to the fifth degree fitted to series of observations $y_1, y_2, \ldots\ldots\ldots, y_n$ corresponding to the values of the independent variate $x_1, x_2, \ldots\ldots\ldots, x_n$ varying by equal intervals.

Writing the polynomials up to the fifth degree in the form,

$$
\begin{aligned}
&y = a_{00}\\
&y = a_{01} + a_{11}x\\
(1)\qquad &y = a_{02} + a_{12}x + a_{22}x^2\\
&y = a_{03} + a_{13}x + a_{23}x^2 + a_{33}x^3\\
&y = a_{04} + a_{14}x + a_{24}x^2 + a_{34}x^3 + a_{44}x^4\\
&y = a_{05} + a_{15}x + a_{25}x^2 + a_{35}x^3 + a_{45}x^4 + a_{55}x^5,
\end{aligned}
$$

it will be seen that there are in all 21 coefficients $a_{00}, a_{01}, \ldots\ldots, a_{55}$, in the set of polynomials up to the fifth degree. For a set of polynomials up to the rth degree there will be $\frac{(r+1)(r+2)}{2}$ coefficients. But not all of these coefficients are independent as some of them repeat themselves and a few others are connected by simple relationships with others, and so, for a complete determination of the least squares solutions of all the polynomials it is not necessary to determine all the coefficients independently. Kerawala has shown that 12 different values are sufficient to determine completely all the 21 constants for the set of polynomials up to the 5th degree. He has also computed tables to facilitate the computation of the 12 values for series up to $n=30$.

It may however be pointed out that by using the orthogonalised terms for the polynomials instead of the simple power terms as in (I) above, the number of independent values that are necessary to determine completely the least squares solutions of all constants of polynomials up to the rth degree can be reduced still further to only $(r+1)$. Six values will thus enable us to determine all the 21 coefficients required for polynomials up to the 5th degree. Tables (Fisher and Yates, 1938 pp. 54-60) are also available for facilitating the determination of the six constants for values of n up to 52.

Assuming the polynomials to be fitted is of the rth degree it can be chosen in the form

$$\text{(II)} \qquad y = A_0\phi_0(x) + A_1\phi_1(x) + A_2\phi_2(x) + \ldots\ldots\ldots + A_r\phi_r(x)$$

where $\phi_r(x)$ is a function of the rth power in x, the independent variate. For simplicity we shall assume here that x is measured in unit intervals from its mid-point so that for n values it will vary

$$\text{(1) from} -\frac{n-1}{2} \text{ to } +\frac{n-1}{2}, \text{ if } n \text{ is odd.}$$

$$\text{(2) from} -n/2 \text{ to } +n/2, \quad \text{if } n \text{ is even.}$$

Now if $\phi_0(x)$, $\phi_1(x)$,........., etc. be chosen to satisfy the condition

$$\text{(III)} \qquad S\{\phi_r(x)\phi_s(x)\}=0 \qquad \text{when } r\neq s$$

the summation extending over the range of the variate x, the least squares solutions of the constants are very easy to determine. The polynomials $\phi_0(x)$, $\phi_1(x)$,... are known as Tchebycheff's polynomials. The general form of such polynomials is connected by the reduction formula (Allan, 1930, p. 312).

$$\text{(IV)} \qquad \phi_r(x)=2(2r-1)x\ \phi_{r-1}(x)-(r-1)^2\{n^2-(r-1)^2\}\phi_{r-2}(x)$$

and their expressions for $r=0, 1, \ldots\ldots, 10$, are given by Allan (1930, p. 319). It will be seen that these are the same as the orthogonal polynomials mentioned by Kerawala as Condon's Polynomials in p. 250 of his paper quoted above. An arithmetical procedure for fitting these polynomials to a set of n observations has been developed by Fisher (Fisher, 1936, p. 156).

The least squares solutions of the constants A_0, A_1,......, A_r are given by

$$\text{(V)} \qquad A_p=\frac{S\{y\phi_p(x)\}}{S\{\phi_p(x)\}^2} \quad \ldots\ldots\text{(Allan, 1930, p. 311).}$$

Using Kerawala's notation that

$$\eta_r=S(yx^r)$$

the least square solutions of the constants up to A_5 are as shown below (Fisher, 1936, p. 150).

$$A_0 = \eta_0 / n$$

$$A_1 = \frac{12}{n(n^2-1)} \eta_1$$

$$A_2 = \frac{180}{n(n^2-1)(n^2-4)} \left(\eta_2 - \frac{n^2-1}{12} \eta_0\right)$$

$$\text{(VI)} \qquad A_3 = \frac{2800}{n(n^2-1)\ldots(n^2-9)} \left(\eta_3 - \frac{3n^2-7}{20} \eta_1\right)$$

$$A_4 = \frac{44100}{n(n^2-1)\ldots(n^2-16)} \left(\eta_4 - \frac{3n^2-13}{14} \eta_2 + \frac{3(n^2-1)(n^2-9)}{560} \eta_0\right)$$

$$A_5 = \frac{698544}{n(n^2-1)\ldots(n^2-25)} \left(\eta_5 - \frac{5(n^2-7)}{18} \eta_3 + \frac{15n^4-230n^2+407}{1008} \eta_1\right).$$

It may be interesting to note that these solutions exactly correspond to the values of a_{00}, a_{11}, a_{22}, a_{33}, a_{44} and a_{55} respectively (Kerawala, 1941, p. 274) in the expressions (I) above and these six values are sufficient to determine all the 21 coefficients involved therein. For, the orthogonal property of the ϕ's used make the least squares solutions of A_0, ..., A_5 independent of each other and hence the process of increasing the degree of curve fitted simply amounts to the addition of one more term to the expression for the previous degree. Thus, if a parabola is already fitted to a series of observations and is given by

$$y = A_0 \phi_0(x) + A_1\phi_1(x) + A_2\phi_2(x),$$

then the cubic curve fitted to the same set of observations is obtained by simply adding $A_3\phi_3(x)$ to the right-hand side of the above equation. Now, if we remember that the least squares solution of a polynomial fitted to a series of observations should be the same whatever may be the form of the polynomial used, it is seen that a_{0r}, a_{1r},......, a_{rr} the least squares solutions of a polynomial of the rth degree should be equal to the coefficients of various powers of x in

$$A_0\phi_0(x) + A_1\phi_1(x) + \ldots\ldots + A_r\phi_r(x).$$

Writing the expressions $\phi_0(x)$, $\phi_1(x)$,......, etc. in full for each degree of the curve fitted and collecting the coefficients of various powers of x, we have, for polynomials up to the 5th degree:

$$y = A_0 = \eta_0/n \qquad \therefore \quad A_0 = a_{00} \qquad \ldots \quad (1)$$

$$y = A_0 + A_1 x \qquad \therefore \quad a_{01} = A_0\,;\; a_{11} = A_1 \qquad \ldots \quad (2)$$

$$y = A_0 + A_1 x + A_2\left(x^2 - \frac{n^2-1}{12}\right) \qquad \ldots \quad (3)$$

$$= \left(A_0 - \frac{n^2-1}{12} A_2\right) + A_1 x + A_2 x^2.$$

(VII) $\therefore\quad a_{02}=A_0-\frac{n^2-1}{12}A_2;$

$a_{12}=A_1$ and $a_{22}=A_2.$

$$y=A_0+A_1x+A_2\left(x^2-\frac{n^2-1}{12}\right)+A_3\left(x^3-\frac{3n^2-7}{20}x\right) \quad \ldots \quad (4)$$

$$=\left(A_0-\frac{n^2-1}{12}A_2\right)+\left(A_1-\frac{3n^2-7}{20}A_3\right)x+A_2x^2+A_3x^3$$

$\therefore\quad a_{03}=A_0-\frac{n^2-1}{12}A_2,\quad a_{13}=A_1-\frac{3n^2-7}{20}A_3;$

$a_{23}=A_2;\quad a_{33}=A_3.$

$$y=A_0+A_1x+A_2\left(x^2-\frac{n^2-1}{12}\right)+A_3\left(x^3-\frac{3n^2-7}{20}x\right)$$
$$+A_4\left(x^4-\frac{3n^2-13}{14}x^2+\frac{3(n^2-1)(n^2-9)}{560}\right) \quad \ldots \quad (5)$$

$\therefore\quad a_{04}=A_0-\frac{n^2-1}{12}A_2+\frac{3(n^2-1)(n^2-9)}{560}A_4;$

$a_{14}=A_1-\frac{3n^2-7}{20}A_3;$

$a_{24}=A_2-\frac{3n^2-13}{14}A_4;$

$a_{34}=A_3;\quad a_{44}=A_4.$

$$y=A_0+A_1x+A_2\left(x^2-\frac{n^2-1}{12}\right)+A_3\left(x^3-\frac{3n^2-7}{20}x\right) \quad \ldots \quad (6)$$
$$+A_4\left(x^4-\frac{3n^2-13}{14}x^2+\frac{3(n^2-1)(n^2-9)}{560}\right)$$
$$+A_5\left(x^5-\frac{5(n^2-7)}{18}x^3+\frac{15n^4-230n^2+407}{1008}x\right)$$

$\therefore\quad a_{05}=A_0-\frac{n^2-1}{12}A_2+\frac{3(n^2-1)(n^2-9)}{560}A_4;$

$a_{15}=A_1-\frac{3n^2-7}{20}A_3+\frac{15n^4-230n^2+407}{1008}A_5;$

$a_{25}=A_2-\frac{3n^2-13}{14}A_4;\quad a_{35}=A_3-\frac{5(n^2-7)}{18}A_5;$

$a_{45}=A_4;\quad a_{55}=A_5.$

The above equations bring out the connection between the least squares solutions of orthogonalised terms and ordinary terms in polynomials up to the fifth degree fitted to a series of n observations. All the equalities given by Kerawala on p. 250 among the 21 coefficients also follow from the above relationships. Hence it is clearly seen that the six coefficients $A_0, A_1, A_2, \ldots, A_5$ are sufficient, if determined, to find the least squares solutions of the 21 coefficients of expressions (I). It is also evident that in general the $(r+1)$ coefficients $A_0, A_1, \ldots, A_r$ will be sufficient to determine all the $\frac{(r+1)(r+2)}{2}$ coefficients in a set of polynomials up to the rth degree.

These orthogonal polynomials are well-known among statisticians and have been used widely by statistical workers for curve-fitting because of the elegant arithmetical procedure for fitting developed by Fisher (4, pp. 151-56). Tables of the functions up to $n=52$ have also been published since, to facilitate rapid computation of A_0 etc., for series of values up to $n=52$, by Fisher and Yates.[2] It will also be seen that the values of $\xi_1, \xi_2, \xi_3, \ldots, \xi_5$ given in these tables agree with the values given for $a_{11}, a_{22}, a_{33}, \ldots, a_{55}$ of Kerawala's tables up to $n=30$, as they should.

The more obvious advantage of using the orthogonalised terms in the polynomial worth mentioning here is the ease with which the sum of squares of residuals can be computed if it is desired to test the goodness of fit. It is not proposed to prove here the results involved, though simple, for which a reference may be made to Allan (1930). It has been shown therein that sum of squares of the residuals after fitting a curve of rth degree is given by

$$\text{(VIII)} \qquad S(y^2) - S\{A_0\phi_0(x)\}^2 - S\{A_1\phi_1(x)\}^2 - \ldots\ldots - S\{A_r\phi_r(x)\}^2$$

where S denotes the summation over the various points of the series. The form at once makes it clear that for each degree of the curve fitted the reduction in the sum of squares is given by the expression

$$A^2S\{\phi(x)\}^2$$

and this enables us also to allocate the total reduction in sum of squares to various individual degrees of the curve fitted with one degree of freedom for each in the analysis of variance. But for the orthogonal property of the terms involved it would not be possible to do so without laborious computations.

One more use of the orthogonalised terms may also be mentioned here. As the coefficients $A_0, A_1, \ldots, A_r$ are independent of one another, they yield a series of independent constants that usefully serve to represent main features of the sequence as indicated by Professor Fisher. He (Fisher, 1936) has suggested that for any series, a', b', c', d', e', f' obtained the following formulae may be taken as a set of independent constants to represent the general distribution features of the series and used in regressional work. These constants have been found very useful in practice for studying distribution features of a variate and the regression functions obtained by correlating these constants with a dependent variate have shown very interesting results (Fisher, 1923, and Kalamkar *et al*, 1940).

$$
\text{(IX)} \quad
\begin{aligned}
A_0 &= a' \\
A_1 &= \frac{6}{n-1} b' \\
A_2 &= \frac{30}{(n-1)(n-2)} c' \\
A_3 &= \frac{140}{(n-1)(n-2)(n-3)} d' \\
A_4 &= \frac{630}{(n-1)(n-2)\ldots(n-4)} e' \\
A_5 &= \frac{2772}{(n-1)(n-2)\ldots(n-5)} f'.
\end{aligned}
$$

It is, therefore, advantageous in many ways to use the orthogonalised terms in a polynomial while fitting it to a series rather than using the ordinary power terms. In view of the enormous amount of arithmetical labour saved if tables are available the tables which are already available up to $n=52$ have now been extended up to $n=75$ by the author of this note and will be published shortly.

Kerawala's paper has, however, thrown much light on the relationships between the least squares solutions of coefficients of ordinary power terms and orthogonalised terms in a polynomial fitted to a series.

METEOROLOGICAL OFFICE,
POONA 5.

REFERENCES

Allan, F. E., 1930, *Proc. Roy. Soc. Edin.*, **50**, Part III, No. 24.

Fisher, R. A., 1923, *Phil. Trans. Roy. Soc. Lond. B.*, **213**, 89.

Fisher, R. A., 1936, *Statistical Methods for Research Workers* (Oliver and Boyd London), 6th edition.

Fisher, R. A., and Yates, F., 1938, *Statistical Tables*, *Tables XXIII*, 54 (Oliver and Boyd, London).

Kalamkar, R. J., and Satakopan, V., 1940, *Ind. Jour. Agri. Sci.*, **10**, 960.

Kerawala, S. M., 1941, *Ind. J. Phys.*, **24**, 241.